

## 第二章-光电转换与 HDR 图像显示

将自然界中的真实场景转换为屏幕上显示出来的图像，往往需要经过两个主要的步骤：第一个是通过摄影设备，将外界的光信息转换为图像信息存储起来，本质上是存储为数字信号；第二个是通过显示设备，将图像信息转换为屏幕输出的光信息。下图展示了从拍摄到现实的电视信号链。在整个过程中，信息流要经过两个重要的非线性映射，才能形成我们在显示设备上看到的图像。其中，相机的非线性映射通常称为光电转换函数（OETF），而显示端的显示器的非线性映射通常称为电光转换函数（EOTF）。通常，OETF 和 EOTF 并不是互逆关系。显示端的 EOTF 往往还包括代表创作者渲染意图的光光转换函数（OOTF），其对观看者在较暗环境下在显示器上观看带来的心理视觉影响进行了修正。

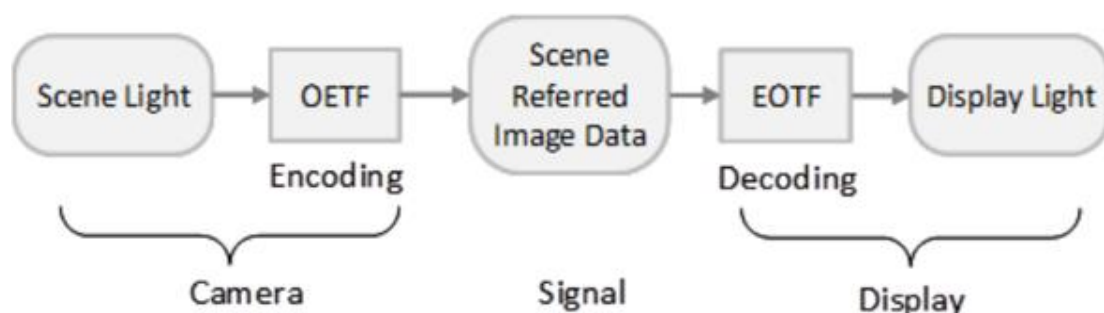


图 1：视频制作与显示流程

在图像获取，存储与显示操作的发展过程中，前面提到的光电转换函数与电光转换函数也经历了多次的发展。

在整个发展过程中，最具有指导意义的是 Gamma 校正，这是一种在传统的 LDR 显示设备上被广泛使用的转换函数。这一章，我们将从 Gamma 校正开始，通过对 Gamma 校正的介绍，解释清楚光电转换与电光转换函数的理论依据与在实际场景中是如何应用的。最后，我们再把目光转移到 HDR 领域上，来介绍在 HDR 领域中，如何设计新的光电转换函数与电光转换函数，来实现 HDR 图像的显示。

### 2.1 Gamma 校正

在最前面必须要强调的是，视觉是人类的主观感受，因此，人眼看到的，感受到的，和自然界中的实际情况与场景可能存在误差。一个最简单的例子就是人眼感受到的，均匀变化的亮度，在自然界中的实际亮度却并不是均匀变化的。这种视觉特性，就是 Gamma 校正产生的原因。人们需要对自然界的亮度进行非线性的映射，才能使人类感受到自然界中的实际景象。

#### 2.1.1 韦伯定律

在详细介绍 Gamma 校正之前，需要先简单了解一个概念：韦伯定律。韦伯定律实际上是一个心理学上的定律。

韦伯定律是用来描述外界刺激的改变与人类感知的改变之间关系的定律，它的公式如下：

$$\Delta\Phi/\Phi = K$$

其中 $\Delta\Phi$ 表示此时的差别阈值，是指人类能够感觉到外界刺激出现改变所需的最小改变量。 $\Phi$ 表示原刺激量， $K$ 为常数。

简而言之，韦伯定律告诉我们，差别阈值与原刺激量成正比关系。也就是说，外界刺激本身越大，差别阈值就越大。

仔细分析一下韦伯定律能够从中发现，人类对于外界刺激的感知并不是线性的。并非当外界刺激强度为 1 时，人类感觉到的强度是 1；外界刺激强度为 2 时，人类感觉到的强度为 2；外界刺激强度为 3 时，人类感觉到的强度就为 3。实际上，人类的感知与外界的实际刺激之间是一种非线性的关系。从韦伯定律的式子中我们可以发现，当原刺激越大时，由于比例关系，差别阈值也就越大。也就是说，当外界的刺激强度越大时，想让人类能够感知到外界的刺激发生了变化的所需变化量，就越大。所以人类的感知从 1 到 2 再到 3 时，所需的变化量是倍数关系的增加，可能外界刺激强度为 1 时，人类感知为 1，刺激强度为 3 时，人类感知为 2，刺激强度为 7 时，人类感知到的强度才为 3。

举一个现实中的例子就能够更好的理解。在一个黑暗的房间中，点燃一根蜡烛，将让人感知到巨大的亮度改变。而如果房间中有 99 根蜡烛，再点燃一根蜡烛人类可能根本感觉不到有所差别。所以，如果想要让人类感知到同样程度的亮度改变，所需的蜡烛数量可能是成百上千根才行。

## 2.1.2 Gamma 校正

根据韦伯定律，我们能够知道人类对于外界刺激的感知是非线性的，人眼对于亮度的感知由此可知，也是非线性的。下面，我们首先介绍人眼对于亮度的实际感受情况，然后以此为基础，介绍 Gamma 校正的相关内容。

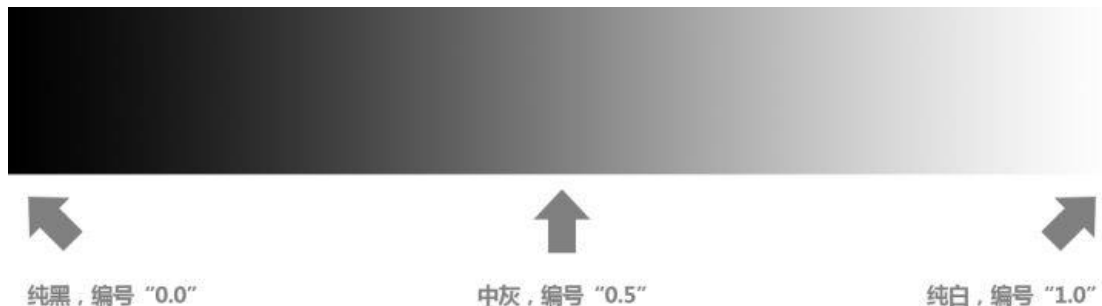


图 2：人眼感受到的亮度

经过一系列实验的测定，人们发现人类能够感知到的连续亮度谱如上图所示。也就是说，如果比上图更加精细，则没有必要，因为人眼无法识别；而如果比上图粗糙，则人眼能够感觉到间断。接着，将最左边纯黑编号为“0.0”，最右端纯白编号为“1.0”，中间定义为中灰“0.5”。这就形成了人眼的亮度谱。

但是，上图中的亮度在自然界中并不是同样成比例的，例如图中的中灰，对应自然界中的亮度是纯白的 20%。

将人眼的亮度谱与自然界中的实际亮度进行建模，就可以得到下面的图像：

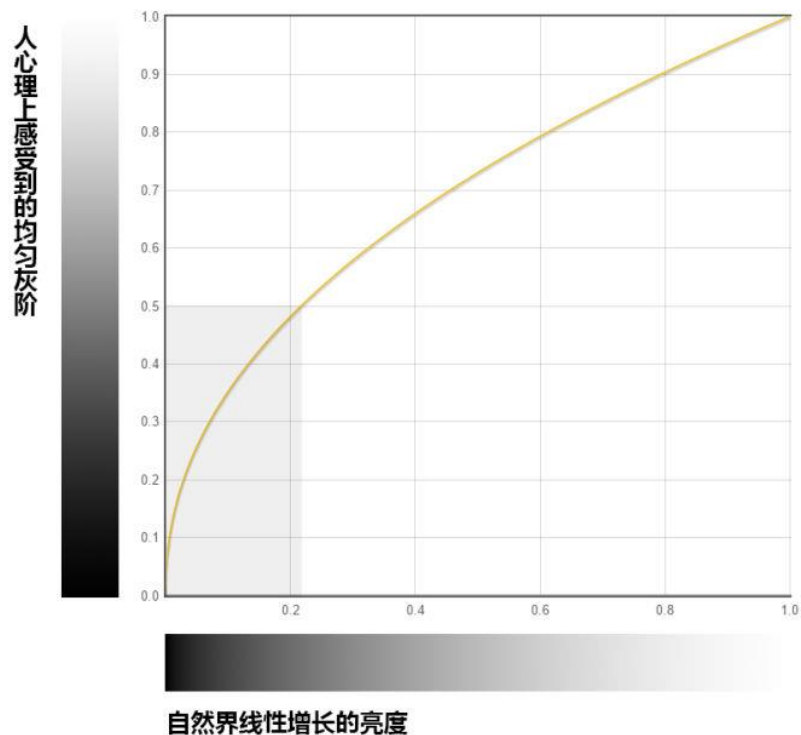


图 3：自然界与人眼感知的亮度对应

从上图可以发现，纵轴上是人眼感受的均匀灰阶，而横轴是自然界中线性增长的亮度，它们之间是一种非线性映射的关系。在低亮度区域，人眼的敏感性较高，因此自然界中很小的亮度范围，都能够被人眼察觉；在高亮度范围，人眼的敏感度较低，因此自然界中的大范围的亮度变化，在人眼感知上却并不大。通过经验以及实验，人们发现可以使用幂函数来拟合这种非线性的映射，用公式表示如下：

$$Y = X^\gamma$$

其中  $X$  是使用 0-1 之间的实数表示的输入亮度， $Y$  是使用 0-1 之间实数表示的输出亮度， $\gamma$  为认为设定的参数，不同的  $\gamma$  选择会产生不同的效果。这种通过调整  $\gamma$  的取值，来调整输入输出亮度之间映射关系的方式，我们称之为 gamma 校正。

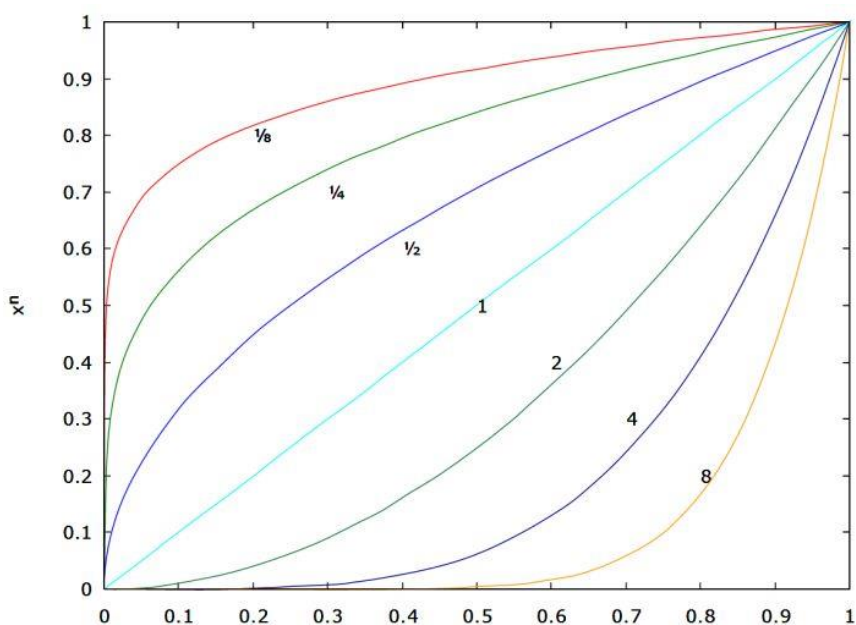


图 4: 不同 $\gamma$ 取值下, 幂函数的图像

根据 $\gamma$ 取值的不同, 通过上图, 我们能够发现 gamma 校正的不同效果:

1. 当 $\gamma < 1$ 时, 曲线上凸, 输出值大于输入值。
2. 当 $\gamma = 1$ 时, 曲线为一条直线, 输入与输出相等, 不进行校正
3. 当 $\gamma > 1$ 时, 曲线下凸, 输出值小于输入值

### 2.1.3 Gamma 校正与图像显示

在传统的图像获取与显示过程中, 将自然界中的真实场景转换为屏幕上显示出来的图像, 往往需要经过两个主要的步骤: 第一个是通过摄影设备, 将外界的光信息转换为三维数组存储起来, 也就是存储为图像信息, 这一过程如前所述, 需要通过光电转换函数进行非线性映射; 第二个是通过显示设备, 将图像信息转换为屏幕输出的光信息, 同样, 这一过程需要电光转换函数的参与。在这两个阶段中, 光电转换函数与电光转换函数实际上都是幂函数, 也就是都进行 Gamma 校正。

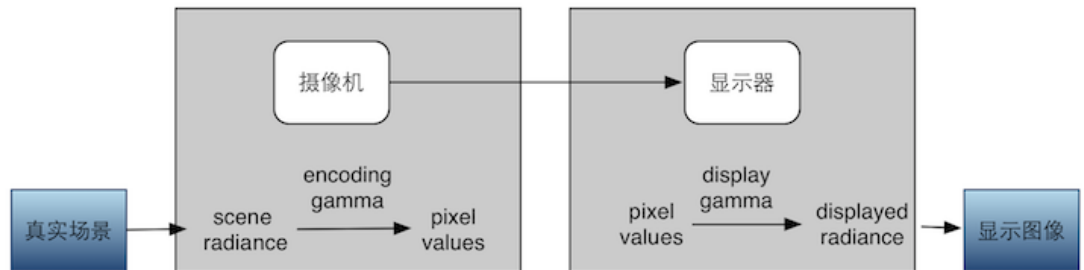


图 5: 真实场景到显示图像的两个转换阶段

首先, 因为图像中的每个像素的取值在 0-255 之间, 因此先进行归一化, 将 0-255 这 256 个数线性映射到 0-1 之间。

接着, 在图像获取阶段, 摄像机中会使用一种 encoding gamma 函数进行 gamma 校正, 这里的 gamma 值小于 1。自然界中的亮度, 首先在摄像机中进行归一化, 映射到 0-1 之间, 接着, 将这个经过了归一化的值输入 encoding gamma 函数中。最后将 encoding gamma 函数的输出值与 0-255 的像素值归一化结果进行比较, 就能得到这个像素点的像素值。

显示阶段的 gamma 校正通常使用一种 display gamma 函数来完成, 这里的 gamma 值大于 1。显示设备读取像素值后, 进行归一化, 然后将归一化结果输入 display gamma 函数中。将输出的值对应的自然光线进行输出, 就是显示阶段做的工作。

理论上来说, 上述两个阶段的 gamma 值乘积应该为 1。因为两个 gamma 校正实际上做的是恰好相反的操作, 一个是将自然光线存储为数字信息, 另一个是将数字信息转换为光线。而为了使自然中的光线与显示设备产生的光线一致, 显然这里的 gamma 值的乘积需要为 1。但是, 在实际操作中, 这两个 gamma 值的乘积并不为 1, 原因有以下两点: 1) 对于传统显示设备来说, 屏幕能够显示的亮度范围与颜色范围与实际情况相差好几个数量级; 2) 在真实的场景中, 原始的场景填充了观察者的所有视野, 而显示设备的亮度往往只局限在一个被周围环境包围的屏幕上。这两个差别使得感知对比度相较于原始场景明显下降了。

为了中和这种现象, 所以我们需要的保证显示的亮度结果在感知上和原始场景是一致的两个 gamma 的乘积并不是 1, 而是经过试验得到的其他值。在电影院这种漆黑的环

境中，这个乘积为 1.5，在明亮的室内这个值为 1.125。个人电脑使用的一个标准叫 sRGB，它使用的 encoding gamma 大约是 0.45（也就是  $1/2.2$ ）。这个值就是为了配合 display gamma 为 2.5 的设备工作的。这样，end-to-end gamma 就是  $0.45 * 2.5 = 1.125$  了。

### 2.1.4 Gamma 校正的意义

按照上一节所说，两段 gamma 校正实际上是一对逆过程，一个将光信息转化为数字信息，一个将数字信息转化为光信息。因此，按照常理来说，并不需要进行这样复杂的非线性映射，只需要将归一化之后的光信息记录下来，然后再让显示设备进行输出即可。从这种角度来说，gamma 校正是一个不必要的操作。

这是一个合理的质疑。问题的根源是在存储的位宽上，如果存储位宽足够大，显然可以使用上述归一化记录方式进行处理，但是，实际生产与应用中的位宽是有限的。在传统的图像存储与显示标准中，一个像素点有三个通道，而每个通道只有 8 位。这也就意味着，图像的灰阶极其有限，与真实世界相比，差了很多个数量级。因此，想要在有限的灰阶上表现出近似于自然界的亮度，就需要对灰阶的分配进行设计，这也是 Gamma 校正的意义。

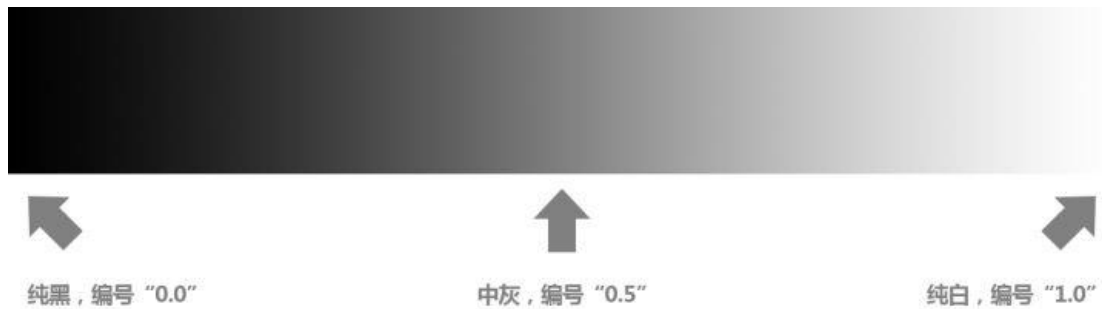


图 6：人眼感受到的亮度

由于人类感知的“中灰”，在实际的自然光线中对应于白光的 20% 的亮度。因此，如果完全按照自然光线的数值进行灰度处理，纯黑到中灰这一段在人类感知中占到一半的灰度，将只能由 20% 的灰阶进行表示，而剩下的一半会有 80% 的灰阶进行表示，这显然是不合理的。Gamma 校正就可以解决这个问题。

在实际操作中，中灰对应的自然光线，显然会被映射为 0.5 的取值，也就对应 128 这一灰阶。这样的结果就是高亮度与低亮度部分分别占有一半的灰阶，这就会使得显示的图像的直观感受更接近于现实。

因此，gamma 校正实际上是用来解决显示设备亮度灰阶不足的问题的。如果将图像存储的编码位宽扩大到 32 位，也就是能完全包含自然界亮度范围的时候，自然就不再需要进行 gamma 校正了，这种情况下，只需要将自然光线的亮度值记录下来，然后对应输出即可。但是，使用 32 位位宽来存储图像过于“奢侈”，因此 gamma 校正在实际生活中起到了广泛而巨大的作用。

## 2.2 HDR 技术中的光电转换

前面提到的传统的 Gamma 校正方式，是针对传统 CRT 显示设备与 LDR 图像设计的，对应的标准是 BT. 1886 标准。而 BT. 1886 标准规定的峰值亮度仅为 100nits，它与传

统 CRT 设备的显示能力相一致。目前的大多数显示设备的显示能力早已突破 100nits 的限制，普通的消费级显示器的亮度一般在 200nits 至 500nits，商业级的监视器亮度可以达到 1000nits 至 2000nits，实验室级别的监视器甚至可以高达 4000nits 至 20000nits。

同时，现代数字运动图像传感器能够捕获动态范围到 80dB 或者比特深度达到 14bit 的线性视频信号。由于实践和历史原因，大多数专业视频制作系统都采用 10 比特，而消费者级别的视频和计算机图形学通常采用 8 比特。在 HDR 技术中，通常采用 10 比特或 12 比特。

由于显示设备的亮度范围得到了提升，同时图像的编码深度也得到了提升，因此传统的 Gamma 校正显然已经不适用于 HDR 技术的光电转换过程。如果强行使用 Gamma 校正进行转换，会在图像与视频中出现明显的分块与分环的情况。

针对上述问题，为了提高显示效果，不同的组织采用了不同的方式进行处理，其中最重要的两种光电转换曲线是：PQ 曲线以及 HLG 曲线。

### 2.2.1 PQ 曲线

为了研究人类视觉系统及喜好，杜比进行了一系列的用户实验，以确定新的 HDR 技术标准以及新的光电转换曲线。

第一个实验室确定新的显示亮度范围。该实验使用一个 2K 的数字电影放映机投影到一个 21 英寸的单色 LCD 面板上，形成一个高对比度的显示器，峰值亮度可以达到 20000 尼特，最小黑电平可以达到 0.004 尼特，色域为 DCI-P3。测试序列共有三个序列集：一个用于检测黑色电平，一个用于检测白色电平，还有一个用于检测高亮电平。实验中观察者处在光线较暗的影院环境下（暗室），位于距离显示设备屏幕高度三倍远的地方。实验结果如下所示，根据实验结果，在亮度范围为 0-10000nits 时，各种信号水平的用户喜好度均超过了 84%，因此 Dolby 提出的 PQ EOTF 曲线决定使用 0-10000nits 作为支持的显示亮度范围。

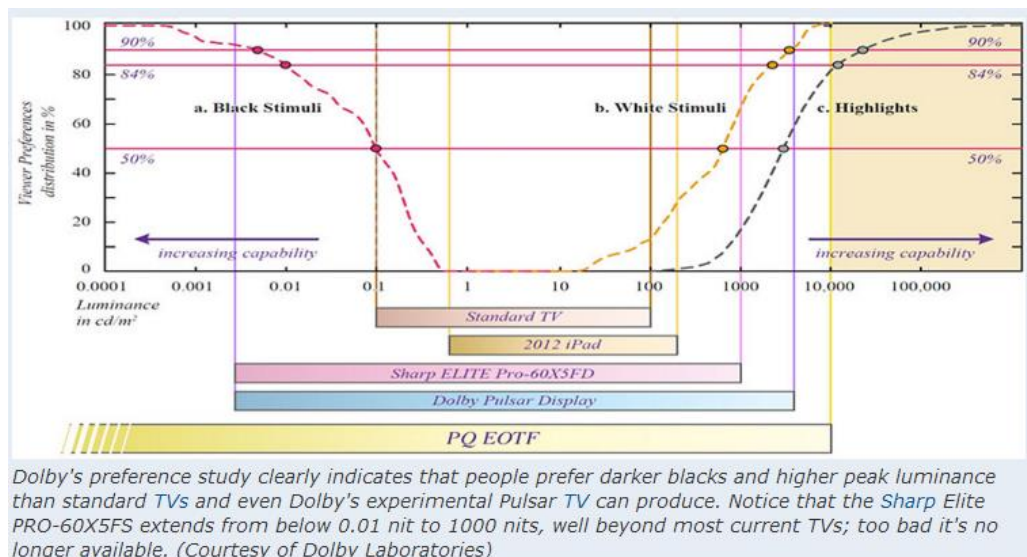


图 7：杜比显示亮度实验结果

由于 8 比特的位深限制了使用更宽广的动态范围，因此需要定义新的 OETF 曲线为视频制作和传输提供更高的动态范围。Dolby 实验室提出了新的 OETF 曲线，称为 PQ 曲



线，目前已经被标准化为 SMPTE ST 2084。

Dolby 提出的 PQ EOTF 的表达式为，它基于 Barten 的 CSF 模型：

$$Y = L \left( \frac{V^{1/m} - c_1}{c_2 - c_3 V^{1/m}} \right)^{1/n}$$

其中：  $0 \leq V \leq 1$ ,  $L = 1000$ ,  $m = \frac{2523}{4096} \times 128 = 78.8438$ ,  $n = \frac{2610}{4096} \times \frac{1}{4} = 0.1593$ ,

$c_1 = \frac{3424}{4096} = 0.8359$ ,  $c_2 = \frac{2413}{4096} \times 32 = 18.8516$ ,  $c_3 = \frac{2392}{4096} \times 32 = 18.6875$ 。

## 10 bit PQ and Rec1886 Curves

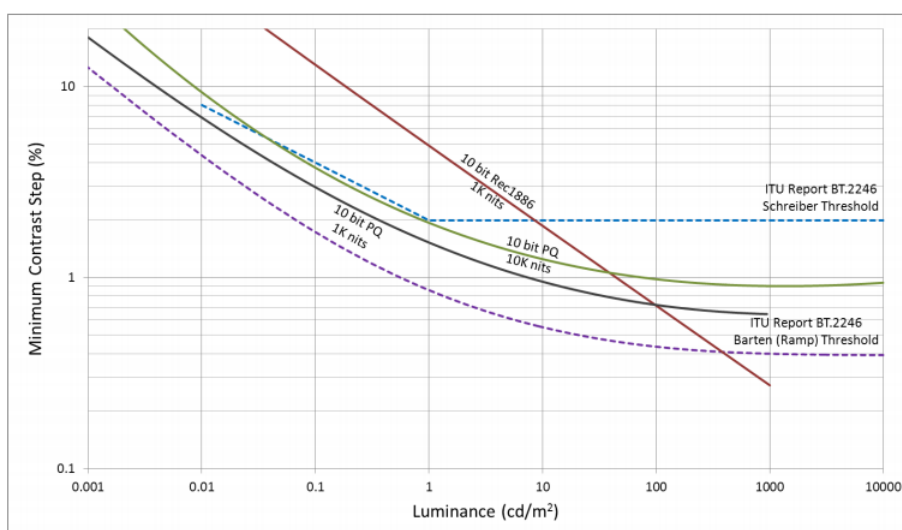


图 8：10 bit PQ 曲线与 Gamma 曲线对比图

## PQ 曲线的优势

Barten Ramp 作为一般图像的共同阈值，大量实验结果显示，Barten 斜坡曲线规定了不同亮度水平下，不出现可见量化误差的最小对比度步长的上界，也就是允许的最大量化步长。当在这条曲线之下时，编码图像可以实现平滑渐变；而当超过这个临界值，也就是在 Barten 斜坡曲线之上时，会出现人眼可察觉的条带效应等量化失真（banding/contouring）。Barten Ramp 曲线表明，在低亮度区域，人眼的视觉灵敏度较低，因此可以分配较少的比特数进行编码；而人眼的视觉灵敏度在高亮度区域非常灵敏，因此编码时需要分配更多的比特数。新的 EOTF 的选择不仅要满足上述特征，还要尽可能的拟合或者贴近 Barten 曲线。这样不仅能够满足人眼视觉特性需求，不会造成量化失真，同时也能够有效利用有限的编码比特数。

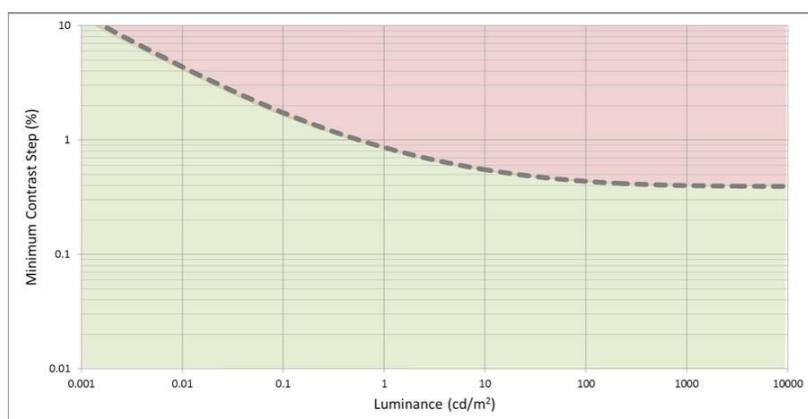


图 9: Barten Ramp 量化误差 当曲线位于虚线上方时, 会产生量化误差, 处在下方时, 不会产生误差。

为使用尽可能少的比特数表示尽可能多的信息, 必须要确定不造成量化误差的最小比特数。如下图所示, 当使用量化精度为 10 比特, 值为 2.4 的 gamma 曲线时, 其低于 10nits 的亮度区域均处于 Barten Ramp 曲线之上, 会造成轻微的量化误差, 并且其峰值亮度最大只能为 100nits。因此, 新的 EOTF 曲线不能使用 10 比特的 gamma 曲线。当使用量化精度为 15nits 的 gamma 曲线时, 整条曲线都在 Barten Ramp 之下, 不会造成量化误差, 峰值亮度可以达到 10000nits。但需要面临的另一个问题是, 15 比特的 gamma 曲线会浪费太多的比特数在高亮度区域。除此之外, 目前的视频传输和处理系统绝大多数都是基于 10 比特进行操作的, 采用 15bitsde gamma 编码不仅带来传输压力, 还需要对市场上大部分的视频传输和制作系统进行更新换代。

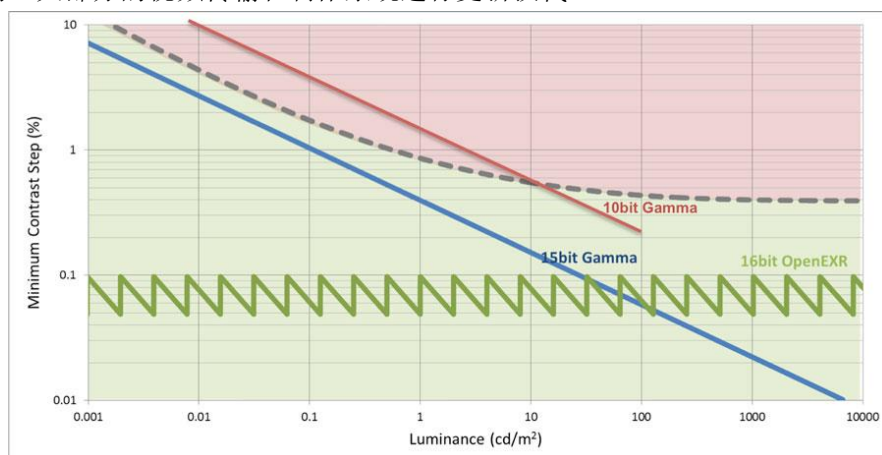


图 10: 曲线对比图 (1)

如下图所示, 如果使用 13 比特的对数曲线时, 与 gamma 曲线相反, 又会在低亮度区域分配更多的比特数。同时可以看到, 用于 HDR 图片存储的 16 比特的 OpenEXR 对应曲线在 Barten 阈值之下, 不会造成可见量化误差。当采用 12 比特的 PQ 曲线作为新的 EOTF 时, 由下图可以看出, 其很好地拟合了 Barten Ramp, 因此不仅符合了人眼视觉系统地特性, 而且充分的利用有限的比特数进行高效合理的编码。



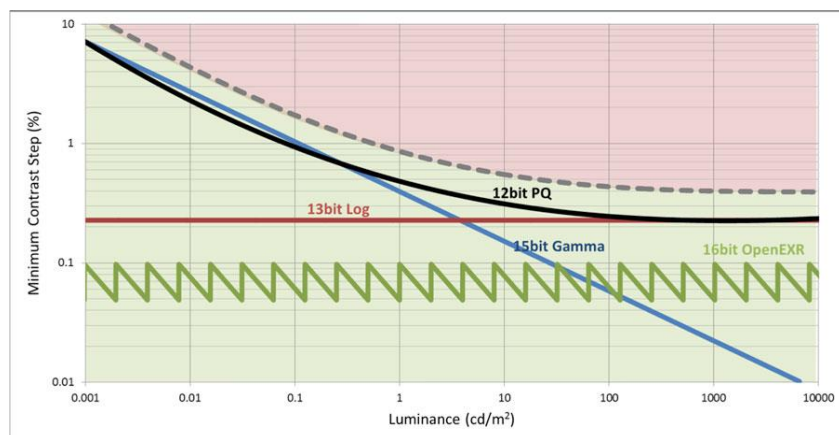
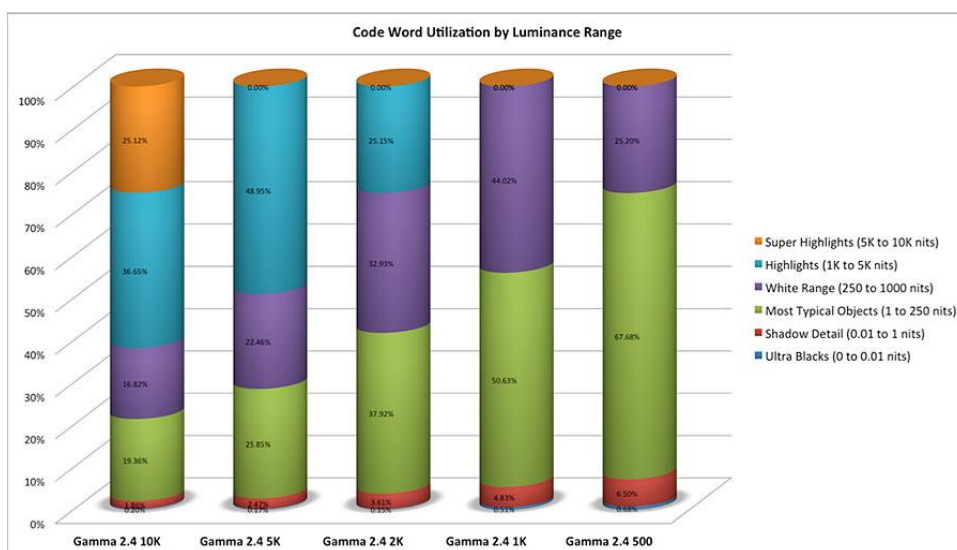


图 10：曲线对比图（2）

下边两张图展示了不同亮度水平下分别采用 Gamma 曲线和 PQ EOTF 进行编码的编码字的利用率。纵轴使用百分比作为度量，100%对应的峰值亮度为 10000nits。Gamma2.4 编码在图像亮度范围为 0-0.01nits 的部分几乎不分配编码字，图像阴影部分分配的编码字也非常少。并且，随着峰值亮度的增加，其在大多数物体所在的亮度范围内以及白光范围内分配的编码字逐渐减少；而使用 PQ 编码，不同峰值亮度水平下很暗的部分（0-0.01nits）以及阴影部分的细节处都分配了一定的编码字。但是在峰值亮度变化时，亮度范围 1-1000nits 的范围分配的编码字差值很小。



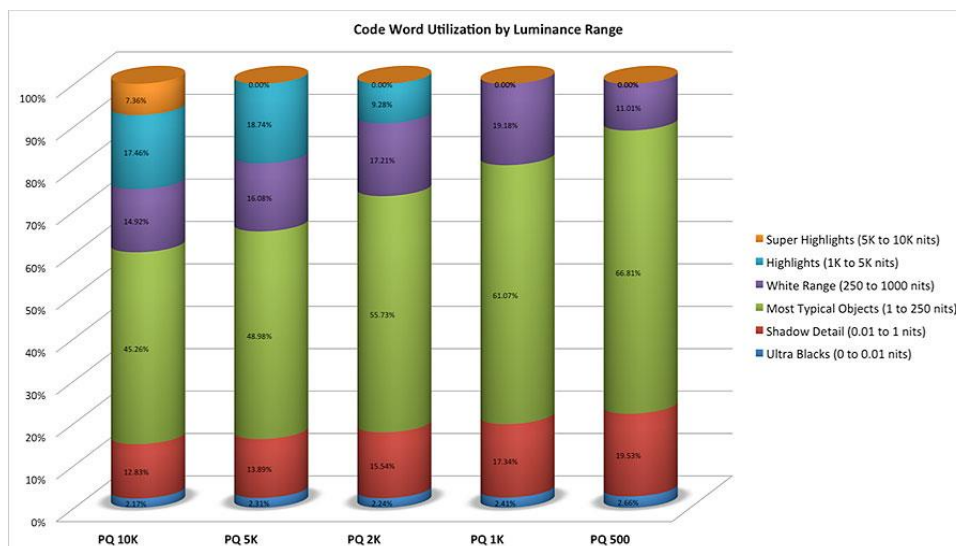


图 11: PQ 曲线与 Gamma 曲线的码字利用率对比

## 2.2.2 HLG 曲线

另外一种与 PQ 相对应的适用于 HDR 的光电转换函数是 HLG (Hybrid-Log Gamma)。与 PQ 的不同在于, PQ 是一种绝对的映射, 亮度值与编码是绝对的一一映射, 而 HLG 是一种相对的映射。

为了增加高光部分的对比度, 通常会对 Rec. 709 OETF 曲线加入一个拐点, 如下图所示。当拐点位置在 85% 的参考白电平, 最大曝光可以改善为原来的 5 倍。这种方法虽然改变了最大曝光, 但是由于拐点位置并没有标准化过, 而且对相机调校的要求很高, 所以这种方法实现起来还是有一定困难。另外, Rec. 709 标准中的 gamma 曲线主要是使噪声在不同的亮度水平下表现一致。

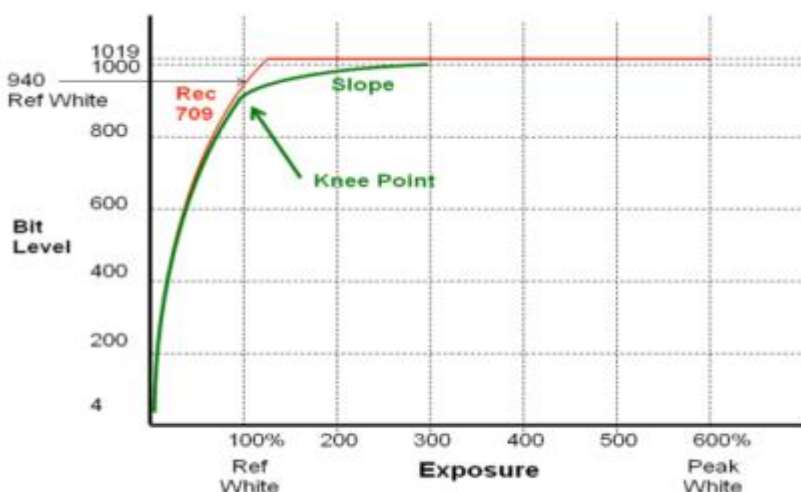


Figure 2: Example of Knee Transfer Characteristic

图 12 拐点图

根据韦伯定律, 图像中的高亮区域的感知量化的阈值基本是恒定的一个常数。因此, 在图像的高亮区域使用对数形式的 OETF 能够利用给定的比特数获得最大的动态范围。

而在图像的暗部，如德维斯-罗斯定律所定义的，当亮度逐渐变小时，可见量化误差会越来越明显。因此，低亮度区域部分的 OETF 可以考虑使用 gamma 曲线，因为 gamma 曲线非常符合德维斯-罗斯定律的约束。综上所述，当动态范围增加时，一个理想的 OETF 应当由两部分组成：低光部分使用传统的 gamma 曲线；而高亮区域使用对数形式的 OETF。这样就形成了新的可用于更高动态范围的 HLG OETF 曲线

$$\text{Weber Fraction} = \frac{1}{N \cdot L} \frac{dL}{dV} = \frac{\gamma}{N} \frac{1}{L^{1/\gamma}}$$

其中 L 为显示亮度，V 为信号电平值，N 为亮化水平（对于 8 比特的视频 N=220；对于 10 比特的视频 N=876）

HLG(Hybrid-Log Gamma)，是由 BBC 和 NHK 提出的一种新的标准，规定了输入端的光电转换特性(OETF)——Hybrid-Log Gamma 曲线，目前已被日本电波产业协会(ARIB)采纳为 ARIB STD-B67 标准。HLG 曲线分为两段，在线性光亮度较小时，其与 SDR 的 gamma 曲线重合，当亮度较高时，其曲线呈现对数特性。在输出端，EOTF 定义为加入了系统 gamma 和亮度调整的 OETF 的逆。支持 10 比特量化和最大亮度 4000nit。

HLG 曲线的数学表达式如下：

$$E' = \begin{cases} r\sqrt{E} & 0 \leq E \leq 1 \\ a \ln(E - b) + c & 1 < E \end{cases}$$

其中，E' 是电平信号，E 表征光照强度。变量 r 为参考电平，一般取 0.5。a、b、c 均为常量，其值分别为 0.17883277、0.28466892、55991073。

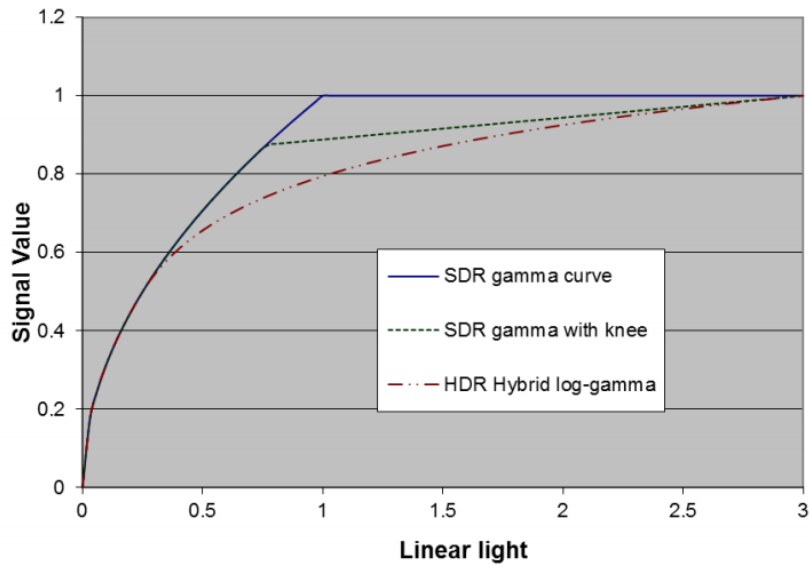


图:13: HLG 曲线与传统 SDR-Gamma 曲线对比 (BBC)

在拍摄端使用 HLG OETF，那么在显示端需要定义对应的 EOTF 才能更好的再现场景。为了很好的适配 HLG OETF，相应的 EOTF 应当满足以下条件：（1）保留创作者的渲染意图；（2）能够适配从黑电平到峰值白电平的整个动态范围；（3）减小量化误差。因此，相应的 EOTF 定义为：

$$Y_d = \alpha Y_s^\gamma + \beta$$

其中  $Y_d$  为显示面板的显示亮度； $Y_s$  表示场景的相对亮度； $\gamma$  表示系统 gamma 值，

取值大小与观看环境相关；而参数  $\alpha$  和  $\beta$  分别控制对比度和亮度，它们共同控制着显

示的峰值亮度和黑电平水平。

$$E = \begin{cases} (E'/r)^2, 0 \leq E' \leq r \\ \exp((E' - c)/a) + b, r < E' \end{cases}$$

根据 HLG OETF 的数学模型可以看出, HLG 的低亮度部分与传统的 gamma 曲线完全吻合, 因此可以不经任何处理直接显示在现有的 SDR 显示器上, 其 10 比特的信号传输制作过程与现有的 SDR 系统完美兼容。另外, 由于不需要元数据, 因此信号是独立于显示器的。